4. Chebyshev's Theorem

The Russian mathematician P.L. Chebyshev (1821-1894) discovered that the fraction of the area between any two values symmetric about the mean is related to the standard deviation.

4. Chebyshev's Theorem

The Russian mathematician P.L. Chebyshev (1821-1894) discovered that the fraction of the area between any two values symmetric about the mean is related to the standard deviation.

Theorem 4.11 (Chebyshev's Theorem) The probability that any random variable X will assume a value within k standard deviations of the mean μ is at least $1 - 1/k^2$. That is,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Chebyshev's Theorem

The theorem gives a conservative estimate of the probability that a random variable assumes a value within k standard deviations of its mean for any real number k.

Chebyshev's Theorem

PROOF (only for continuous case)

By our previous definition of the variance of X we can write

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2]\\ &= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx\\ &= \int_{-\infty}^{\mu-k\sigma}(x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma}(x-\mu)^2 f(x)dx\\ &\quad + \int_{\mu+k\sigma}^{\infty}(x-\mu)^2 f(x)dx\\ &\geq \int_{-\infty}^{\mu-k\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty}(x-\mu)^2 f(x)dx\end{aligned}$$

since the second of the three integrals is nonnegative. Now, since $|x-\mu| \geq k\sigma$ wherever $x \geq \mu + k\sigma$ or $x \leq \mu - k\sigma$, we have $(x-\mu)^2 \geq k^2\sigma^2$ in both remaining integral. It follows that

Chebyshev's Theorem

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)dx$$

and that

$$\frac{1}{k^2} \geq \int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx$$

Hence

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Chebyshev's Theorem

For $k = 2$ the theorem states that $3/4$ or more of the observations of any distribution lie in the interval $\mu \pm 2\sigma$.

Chebyshev's Theorem

For $k = 2$ the theorem states that $3/4$ or more of the observations of any distribution lie in the interval $\mu \pm 2\sigma$.

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{2^2}.$$

Chebyshev's Theorem

For $k = 2$ the theorem states that $3/4$ or more of the observations of any distribution lie in the interval $\mu \pm 2\sigma$.

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{2^2}.$$

Similarly, at least eight-ninths of the observations of any distribution fall in the interval $\mu \pm 3\sigma$.

Chebyshev's Theorem

For $k = 2$ the theorem states that $3/4$ or more of the observations of any distribution lie in the interval $\mu \pm 2\sigma$.

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{2^2}.$$

Similarly, at least eight-ninths of the observations of any distribution fall in the interval $\mu \pm 3\sigma$.

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \geq 1 - \frac{1}{3^2}.$$

Chebyshev's Theorem

Example 4.22 A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find

(a) $P(-4 < X < 20)$, (b) $P(|X - 8| \geq 6)$

Chebyshev's Theorem

Example 4.22 A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find
(a) $P(-4 < X < 20)$, (b) $P(|X - 8| \geq 6)$

Solution......

$P(-4 < X < 20) \geq \frac{15}{16}$, $P(|X - 8| \geq 6) \leq \frac{1}{4}$.

Chebyshev's Theorem

The Chebyshev's Theorem is a distribution-free result.

Therefore, the results are usually weak.

Chebyshev's Theorem

The Chebyshev's Theorem is a distribution-free result.
Therefore, the results are usually weak.

The value given by the theorem is **a lower bound only.**

Chebyshev's Theorem

The Chebyshev's Theorem is a distribution-free result. Therefore, the results are usually weak.

The value given by the theorem is **a lower bound only.**

We know that the probability of a random variable falling within two standard deviations of the mean can be no less than $3/4$, but we never know how much more it might actually be.

Chebyshev's Theorem

The Chebyshev's Theorem is a distribution-free result. Therefore, the results are usually weak.

The value given by the theorem is **a lower bound only.**

We know that the probability of a random variable falling within two standard deviations of the mean can be no less than $3/4$, but we never know how much more it might actually be.

Only when the probability distribution is known can we determine exact probabilities.